



1600

RAW SEQUENCE LISTING

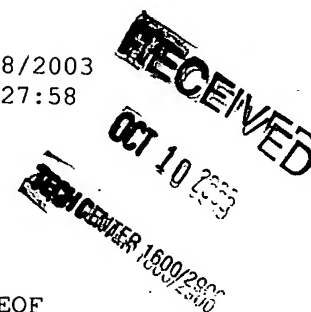
PATENT APPLICATION: US/10/090,455A

DATE: 10/08/2003

TIME: 14:27:58

Input Set : E:\406.app.txt

Output Set: N:\CRF4\10082003\J090455A.raw



ENTERED

4 <110> APPLICANT: Chen, Hongyun
 5 Le Bihan, Stephane
 7 <120> TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
 10 <130> FILE REFERENCE: 100103.406
 12 <140> CURRENT APPLICATION NUMBER: US 10/090,455A
 13 <141> CURRENT FILING DATE: 2002-03-01
 15 <160> NUMBER OF SEQ ID NOS: 23
 17 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 19 <210> SEQ ID NO: 1
 20 <211> LENGTH: 3455
 21 <212> TYPE: DNA
 22 <213> ORGANISM: Homo sapiens
 24 <400> SEQUENCE: 1
 25 gccaccatgg cggagaaggg gctggaggcc gtgggctgtg gactaggggc gggggctgtg 60
 26 gccatggccg tgacgctgga ggacggggcg gaacccccctg tgctgaccac gcacctgaag 120
 27 aaggtggaga accacatcac tgaagcccag cgcttctccc acctgcccac gcgctcagcc 180
 28 gtggacatcg agttcgtgga gctgtcctat tccgtgcggg agggggccctg ctggcgcaaa 240
 29 aggggttata agacccttct caagtgcctc tcaggtaaatt tctgccgcgg ggagctgatt 300
 30 ggcatcatgg gccctcagg ggctggcaag gctgacatgc gtaaatggaa ggccacggga gctgaggacc 360
 31 agggagtctg gaatgaagg gcatatcctg gttaattgaa ggccacggga gctgaggacc 420
 32 ttcgcgaaga tgtcctgcta catcatgcaa gatgacatgc tgctgccgca cctcacgggtg 480
 33 ttggaagcca tgatggtctc tgctaacctg aatcttactg agaattccga tgtgaaaaac 540
 34 gatctcgtga cagagatcct gacggcactg ggctgatgt cgtgctccca cagcaggaca 600
 35 gccctgctct ctggcgggca gaggaagcgt ctggccatcg ccctggagct ggtcaacaac 660
 36 ccgctgttca tgttctttga tgagcccacc agtggctctg atagcgctc ttgtttccaa 720
 37 gtggtgtccc tcatgaagtc cctggcacag gggggccgta ccatcatctg caccatccac 780
 38 cagcccagtg ccaagctctt tgagatgttt gacaagctct acatcctgag ccagggtcag 840
 39 tgcatttca aaggagtggg caccaacctg atccccatc taagaggact cggcttgcag 900
 40 tgccccacct accacaaccc ggctgacttc atcatcgagg tggcctctgg cgagtatgga 960
 41 gacctgaacc ccatgttgtt cagggtctgt cagaatggg tgtgcgctat ggctgagaag 1020
 42 aagagcagcc ctgagaagaa cgaggtccct gccccatgcc ctcttgtcc tccggaagtg 1080
 43 gatcccattg aaagccacac ctttgccacc agcaccctca cacagttctg catcctcttc 1140
 44 aagaggacct tcctgtccat cctcagggac acggtcctga ccacctacg gttcatgtcc 1200
 45 cacgtggtta ttggcggtgt catcgccctc ctctacctgc atattggcga cgatgccagc 1260
 46 aaggtcttca acaacacggg ctgcctcttc ttctccatgc tttctctcat gttcgccgcc 1320
 47 ctcatgccaa ctgtgtcac cttcccctta gagatggcgg tcttcatgag ggagcacctc 1380
 48 aactactggt acagcctcaa agcgtattac ctggccaaga ccatggctga cgtgcccttt 1440
 49 caggtggtgt gtcgggtggt ctactgcagc attgtgtact ggatgacggg ccagcccgct 1500
 50 gagaccagcc gcttctgtct cttctcagcc ctggccaccg ccaccgcctt ggtggcccaa 1560
 51 tctttggggc tgctgatcgg agctgcttcc aactccctac aggtggccac ttttggggc 1620
 52 ccagttaccg ccatccctgt cctcttggtc tccggttct ttgtcagctt caagaccatc 1680
 53 cccacttacc tgcaatggag ctccatctc tcctatgtca ggtatggctt tgagggtgtg 1740
 54 atcctgacga tctatggcat ggagcgagga gacctgacat gtttagagga acgctgccc 1800

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/090,455A

DATE: 10/08/2003

TIME: 14:27:58

Input Set : E:\406.app.txt

Output Set: N:\CRF4\10082003\J090455A.raw

```

55 ttccgggagc cacaagagcat cctccgagcg ctggatgtgg aggatgccaa gctctacatg 1860
56 gacttcctgg tcttggggcat cttcttctcta gccctgcggc tgctggccta ccttgtgctg 1920
57 cgttaccggg tcaagtcaga gagatagagg cttgccccag cctgtacccc agccccctgca 1980
58 gcaggaagcc cccagtcacca gccctttggg actgttttaa ccttatagac ttgggcactg 2040
59 gttcctggcg gggctatcct ctccctccctt ggctcctcca caggctggct gtcggactgc 2100
60 gctcccagcc tgggctctgg gagtgggggc tccagccctc cccactatgc ccaggagtct 2160
61 tcccaagttg atgcggtttg tagcttctct cctactctct ccaacacctg catgcaaaga 2220
62 ctactgggag gctgctgcct ccttctctgcc catggcaccc tcctctgctg tctgcctggg 2280
63 agccctaggc tctctagggc cccacttaca actgacccaa gtggcccccct ctgggggtcc 2340
64 ccaccaacaca agtggtttgta aactgggctg ctataaggtt ggagttccag ggctggggccc 2400
65 tgggtggagtc cactggaagt cccattatgg atgttgaaat ggacagggaa ggactctgga 2460
66 agtctcttcc tctctctcct cttctctcca cccctagacc ctggctgact tggacaatct 2520
67 gccaggacag aagctgggtt ttctgtctag ctcaccactc ccaatcctgg ggattggaga 2580
68 ggctgggggc tgtgggatgc cccatccccc tccccatcac ctttgggtgg ggacgggcct 2640
69 ggtggcacct gtgcaataat gtctgtgttt ctctcccacc tgccactgga actggagaat 2700
70 gcactttatt ctgggcgggg ggtgagtggg ggaagaccca accctccttt ctgctgccc 2760
71 ctaacgcgatg caggtctctg tgatgctccc tccctctccg gagtgacagg cacatacatg 2820
72 agaacaggcc atctcagccc tacacacttg ccatccccta cagcacagag gaagagtgat 2880
73 ggtggcatgc tgggtggggc ggggtgctgg gggaggacag tgccaacctc ctctggggga 2940
74 tcccatgttg gagactctaa ggataaggct ggtgctgccc aggggtgtcta caggaactgc 3000
75 aggtgtctac cccaagtct tccctcctcc caagccaggg gtggcacagg gcactagatc 3060
76 cctggagttc aggaaccaac acaagcaca ccacgggcat aagttggcct tggccactgc 3120
77 caccacggc cctccttttg tgcctcatgc tggcatcttc actcccctac cccttccccca 3180
78 gccactgctg ctctattcaa cttctgtcca tgtccctcca ctgttctat cagcagggtg 3240
79 cccctgggca tcagaacagc ctgccctggg caccagggtg cagacacact cagagcatgt 3300
80 ctggctttcc tgggtgggtcc aggtcattc tgcttctgat ttcccctccc ccagggtc 3360
81 ttttccccct ttttctgta cacatccctg tctacctcct ctacccctgc cacagattct 3420
82 tcctatcaca cagggatgcc agttgtattt gtggg 3455

```

84 <210> SEQ ID NO: 2

85 <211> LENGTH: 646

86 <212> TYPE: PRT

87 <213> ORGANISM: Homo sapiens

89 <400> SEQUENCE: 2

```

90 Met Ala Glu Lys Ala Leu Glu Ala Val Gly Cys Gly Leu Gly Pro Gly
91 1 5 10 15
92 Ala Val Ala Met Ala Val Thr Leu Glu Asp Gly Ala Glu Pro Pro Val
93 20 25 30
94 Leu Thr Thr His Leu Lys Lys Val Glu Asn His Ile Thr Glu Ala Gln
95 35 40 45
96 Arg Phe Ser His Leu Pro Lys Arg Ser Ala Val Asp Ile Glu Phe Val
97 50 55 60
98 Glu Leu Ser Tyr Ser Val Arg Glu Gly Pro Cys Trp Arg Lys Arg Gly
99 65 70 75 80
100 Tyr Lys Thr Leu Leu Lys Cys Leu Ser Gly Lys Phe Cys Arg Arg Glu
101 85 90 95
102 Leu Ile Gly Ile Met Gly Pro Ser Gly Ala Gly Lys Ser Thr Phe Met
103 100 105 110
104 Asn Ile Leu Ala Gly Tyr Arg Glu Ser Gly Met Lys Gly Gln Ile Leu
105 115 120 125

```

RAW SEQUENCE LISTING

DATE: 10/08/2003

PATENT APPLICATION: US/10/090,455A

TIME: 14:27:58

Input Set : E:\406.app.txt

Output Set: N:\CRF4\10082003\J090455A.raw

```

106 Val Asn Gly Arg Pro Arg Glu Leu Arg Thr Phe Arg Lys Met Ser Cys
107      130                      135                      140
108 Tyr Ile Met Gln Asp Asp Met Leu Leu Pro His Leu Thr Val Leu Glu
109 145                      150                      155                      160
110 Ala Met Met Val Ser Ala Asn Leu Asn Leu Thr Glu Asn Pro Asp Val
111                      165                      170                      175
112 Lys Asn Asp Leu Val Thr Glu Ile Leu Thr Ala Leu Gly Leu Met Ser
113                      180                      185                      190
114 Cys Ser His Thr Arg Thr Ala Leu Ser Gly Gly Gln Arg Lys Arg
115                      195                      200                      205
116 Leu Ala Ile Ala Leu Glu Leu Val Asn Asn Pro Pro Val Met Phe Phe
117      210                      215                      220
118 Asp Glu Pro Thr Ser Gly Leu Asp Ser Ala Ser Cys Phe Gln Val Val
119 225                      230                      235                      240
120 Ser Leu Met Lys Ser Leu Ala Gln Gly Gly Arg Thr Ile Ile Cys Thr
121                      245                      250                      255
122 Ile His Gln Pro Ser Ala Lys Leu Phe Glu Met Phe Asp Lys Leu Tyr
123                      260                      265                      270
124 Ile Leu Ser Gln Gly Gln Cys Ile Phe Lys Gly Val Val Thr Asn Leu
125      275                      280                      285
126 Ile Pro Tyr Leu Lys Gly Leu Gly Leu His Cys Pro Thr Tyr His Asn
127      290                      295                      300
128 Pro Ala Asp Phe Ile Ile Glu Val Ala Ser Gly Glu Tyr Gly Asp Leu
129 305                      310                      315                      320
130 Asn Pro Met Leu Phe Arg Ala Val Gln Asn Gly Leu Cys Ala Met Ala
131                      325                      330                      335
132 Glu Lys Lys Ser Ser Pro Glu Lys Asn Glu Val Pro Ala Pro Cys Pro
133                      340                      345                      350
134 Pro Cys Pro Pro Glu Val Asp Pro Ile Glu Ser His Thr Phe Ala Thr
135      355                      360                      365
136 Ser Thr Leu Thr Gln Phe Cys Ile Leu Phe Lys Arg Thr Phe Leu Ser
137      370                      375                      380
138 Ile Leu Arg Asp Thr Val Leu Thr His Leu Arg Phe Met Ser His Val
139 385                      390                      395                      400
140 Val Ile Gly Val Leu Ile Gly Leu Leu Tyr Leu His Ile Gly Asp Asp
141                      405                      410                      415
142 Ala Ser Lys Val Phe Asn Asn Thr Gly Cys Leu Phe Phe Ser Met Leu
143                      420                      425                      430
144 Phe Leu Met Phe Ala Ala Leu Met Pro Thr Val Leu Thr Phe Pro Leu
145      435                      440                      445
146 Glu Met Ala Val Phe Met Arg Glu His Leu Asn Tyr Trp Tyr Ser Leu
147      450                      455                      460
148 Lys Ala Tyr Tyr Leu Ala Lys Thr Met Ala Asp Val Pro Phe Gln Val
149 465                      470                      475                      480
150 Val Cys Pro Val Val Tyr Cys Ser Ile Val Tyr Trp Met Thr Gly Gln
151                      485                      490                      495
152 Pro Ala Glu Thr Ser Arg Phe Leu Leu Phe Ser Ala Leu Ala Thr Ala
153      500                      505                      510
154 Thr Ala Leu Val Ala Gln Ser Leu Gly Leu Leu Ile Gly Ala Ala Ser

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/090,455A

DATE: 10/08/2003

TIME: 14:27:58

Input Set : E:\406.app.txt

Output Set: N:\CRF4\10082003\J090455A.raw

```

155          515          520          525
156 Asn Ser Leu Gln Val Ala Thr Phe Val Gly Pro Val Thr Ala Ile Pro
157          530          535          540
158 Val Leu Leu Phe Ser Gly Phe Phe Val Ser Phe Lys Thr Ile Pro Thr
159 545          550          555          560
160 Tyr Leu Gln Trp Ser Ser Tyr Leu Ser Tyr Val Arg Tyr Gly Phe Glu
161          565          570          575
162 Gly Val Ile Leu Thr Ile Tyr Gly Met Glu Arg Gly Asp Leu Thr Cys
163          580          585          590
164 Leu Glu Glu Arg Cys Pro Phe Arg Glu Pro Gln Ser Ile Leu Arg Ala
165          595          600          605
166 Leu Asp Val Glu Asp Ala Lys Leu Tyr Met Asp Phe Leu Val Leu Gly
167          610          615          620
168 Ile Phe Phe Leu Ala Leu Arg Leu Leu Ala Tyr Leu Val Leu Arg Tyr
169 625          630          635          640
170 Arg Val Lys Ser Glu Arg
171          645
174 <210> SEQ ID NO: 3
175 <211> LENGTH: 1941
176 <212> TYPE: DNA
177 <213> ORGANISM: Homo sapiens
179 <400> SEQUENCE: 3
180 atggcggaga aggcgctgga ggccgtgggc tgtggactag ggccgggggc tgtggccatg 60
181 gccgtgacgc tggaggacgg ggcggaaccc cctgtgctga ccacgcacct gaagaagggtg 120
182 gagaaccaca tcaactgaagc ccagcgcttc tcccacctgc ccaagcgctc agccgtggac 180
183 atcgagttag tggagctgtc ctattccgtg cgggaggggc cctgctggcg caaaaggggt 240
184 tataagaccc ttctcaagtg cctctcaggc aaattctgcc gccgggagct gattggcatc 300
185 atggggccct caggggctgg caagtctaca ttcatgaaca tcttggcagg atacaggagg 360
186 tctggaatga aggggcagat cctgggtaat ggaaggccac gggagctgag gaccttccgc 420
187 aagatgtcct gctacatcat gcaagatgac atgctgctgc cgcaacctac ggtgttgaa 480
188 gccatgatgg tctctgctaa cctgaatctt actgagaatc ccgatgtgaa aaacgatctc 540
189 gtgacagaga tcctgacggc actgggcctg atgtcgtgct cccacacgag gacagccctg 600
190 ctctctggcg ggcagaggaa gcgtctggcc atcgccctgg agctgggtcaa caaccgcct 660
191 gtcatgttct ttgatgagcc caccagtggc ctggatagcg cctcttggtt ccaagtgggtg 720
192 tccctcatga agtccttggc acaggggggc cgtaccatca tctgcaccat ccaccagccc 780
193 agtgccaagc tctttgagat gtttgacaag ctctacatcc tgagccaggg tcagtgcac 840
194 ttcaaaggag tggtcaccaa cctgatcccc tatctaaagg gactcggctt gcattgcccc 900
195 acctaccaca acccggtgga cttcatcatc gaggtggcct ctggcgagta tggagacctg 960
196 aaccccatgt tgttcagggc tgtgcagaat gggctgtgct ctatggctga gaagaagagc 1020
197 agccctgaga agaacgaggt cctgccccca tgccctcctt gtccctccga agtggatccc 1080
198 attgaaagcc acacctttgc caccagcacc ctacacagc tctgcatcct cttcaagagg 1140
199 accttcctgt ccatacctcag ggacacggtc ctgaccacc tacggttcat gtcccacgtg 1200
200 gttattggcg tgtcatcggt cctcctctac ctgcatattg gcgacgatgc cagcaaggtc 1260
201 ttcaacaaca ccggtgcctt cttcttctcc atgctgttcc tcatgttcgc cgccctcatg 1320
202 ccaactgtgc tcaccttccc cttagagatg gcggtcttca tgaggagga cctcaactac 1380
203 tggtagagcc tcaaagcgta ttacctggcc aagaccatgg ctgacgtgcc ctttcagggtg 1440
204 gtgtgtccgg tgggtacttg cagcattgtg tactggatga cgggccagcc cgctgagacc 1500
205 agccgcttcc tgccttcttc agccctggcc accgccaccg ccttgggtggc ccaatctttg 1560
206 gggctgctga tcggagctgc ttccaactcc ctacaggtgg ccacttttgt gggcccagtt 1620

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/090,455A

DATE: 10/08/2003

TIME: 14:27:58

Input Set : E:\406.app.txt

Output Set: N:\CRF4\10082003\J090455A.raw

```

207 accgccatcc ctgtcctctt gttctccggc ttctttgtca gcttcaagac catccccact 1680
208 tacctgcaat ggagctccta tctctcctat gtcaggtagt gctttgaggg tgtgatcctg 1740
209 acgatctatg gcatggagcg aggagacctg acatgttttag aggaacgctg cccgttccgg 1800
210 gagccacaga gcatcctccg agcgctggat gtggaggatg ccaagctcta catggacttc 1860
211 ctgggtcttgg gcatcttctt cctagccctg cggctgctgg cctaccttgt gctgcgttac 1920
212 cgggtcaagt cagagagata g                                     .1941
214 <210> SEQ ID NO: 4
215 <211> LENGTH: 674
216 <212> TYPE: PRT
217 <213> ORGANISM: Homo sapiens
219 <400> SEQUENCE: 4
220 Met Ala Ala Phe Ser Val Gly Thr Ala Met Asn Ala Ser Ser Tyr Ser
221 1 5 10 15
222 Ala Glu Met Thr Glu Pro Lys Ser Val Cys Val Ser Val Asp Glu Val
223 20 25 30
224 Val Ser Ser Asn Met Glu Ala Thr Glu Thr Asp Leu Leu Asn Gly His
225 35 40 45
226 Leu Lys Lys Val Asp Asn Asn Leu Thr Glu Ala Gln Arg Phe Ser Ser
227 50 55 60
228 Leu Pro Arg Arg Ala Ala Val Asn Ile Glu Phe Arg Asp Leu Ser Tyr
229 65 70 75 80
230 Ser Val Pro Glu Gly Pro Trp Trp Arg Lys Lys Gly Tyr Lys Thr Leu
231 85 90 95
232 Leu Lys Gly Ile Ser Gly Lys Phe Asn Ser Gly Glu Leu Val Ala Ile
233 100 105 110
234 Met Gly Pro Ser Gly Ala Gly Lys Ser Thr Leu Met Asn Ile Leu Ala
235 115 120 125
236 Gly Tyr Arg Glu Thr Gly Met Lys Gly Ala Val Leu Ile Asn Gly Leu
237 130 135 140
238 Pro Arg Asp Leu Arg Cys Phe Arg Lys Val Ser Cys Tyr Ile Met Gln
239 145 150 155 160
240 Asp Asp Met Leu Leu Pro His Leu Thr Val Gln Glu Ala Met Met Val
241 165 170 175
242 Ser Ala His Leu Lys Leu Gln Glu Lys Asp Glu Gly Arg Arg Glu Met
243 180 185 190
244 Val Lys Glu Ile Leu Thr Ala Leu Gly Leu Leu Ser Cys Ala Asn Thr
245 195 200 205
246 Arg Thr Gly Ser Leu Ser Gly Gly Gln Arg Lys Arg Leu Ala Ile Ala
247 210 215 220
248 Leu Glu Leu Val Asn Asn Pro Pro Val Met Phe Phe Asp Glu Pro Thr
249 225 230 235 240
250 Ser Gly Leu Asp Ser Ala Ser Cys Phe Gln Val Val Ser Leu Met Lys
251 245 250 255
252 Gly Leu Ala Gln Gly Gly Arg Ser Ile Ile Cys Thr Ile His Gln Pro
253 260 265 270
254 Ser Ala Lys Leu Phe Glu Leu Phe Asp Gln Leu Tyr Val Leu Ser Gln
255 275 280 285
256 Gly Gln Cys Val Tyr Arg Gly Lys Val Cys Asn Leu Val Pro Tyr Leu
257 290 295 300

```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 10/08/2003
PATENT APPLICATION: US/10/090,455A TIME: 14:27:59

Input Set : E:\406.app.txt
Output Set: N:\CRF4\10082003\J090455A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:13; Xaa Pos. 579,598

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/090,455A

DATE: 10/08/2003

TIME: 14:27:59

Input Set : E:\406.app.txt

Output Set: N:\CRF4\10082003\J090455A.raw

L:836 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:13 after pos.:576
M:341 Repeated in SeqNo=13